SYSTEM AND METHOD FOR DETERMINING AFFIXES OF WORDS

ABSTRACT OF THE DISCLOSURE

A computer system and a method for analyzing text in one or more electronic documents are disclosed. The computer system comprises one or more system interfaces; and an affix process that determines one or more affixes of one or more words in one or more of the documents and provides the affixes to the system interface. The preferred embodiment of the invention may be used to build a domain specific morphology lexicon for NLP applications so that they can recognize out-of-vocabulary words. The disclosed procedure utilizes the fact that the processes of discovering prefixes and suffixes are not independent. Many words, especially in technical documents, have complex morphological structures, and thus the knowledge about prefixes helps the discovery of suffixes and vice versa.